RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: _10/518, 019A_
Source: _Pct/lo_
Date Processed by STIC: _9/5/05_

ENTERED

PCT

```
 3 <110> APPLICANT: Commissariat a l'Energie Atomique
 4        Centre National de la Recherche Scientifique
 5        GONDRY, Muriel
 6        GENET, Roger
 7        LAUTRU, Sylvie
 8        PERNODET, Jean-Luc
10 <120> TITLE OF INVENTION: Polynucleotides and polypeptides coded by said
polynucleotides
11        involved in the synthesis of diketopiperazine derivatives
13 <130> FILE REFERENCE: CGA263/83FR
15 <140> CURRENT APPLICATION NUMBER: US/10/518,019A
16 <141> CURRENT FILING DATE: 2004-12-15
18 <160> NUMBER OF SEQ ID NOS: 23
20 <170> SOFTWARE: PatentIn version 3.3
22 <210> SEQ ID NO: 1
23 <211> LENGTH: 657
24 <212> TYPE: DNA
25 <213> ORGANISM: Streptomyces noursei
27 <400> SEQUENCE: 1
28 gtgaggcgcc acccatcgca ttcgccgtac cgcggcgggt gtgaggtgcg cccaaaaaga    60
30 aggggattga tgttagctca cagttcatct gaatcgccgc cggaatcctt gccggacgcg   120
32 tggacggtcc tcaaaacccg taccgccgtc cgcaattacg cgaaagagcc ggtcgacgac   180
34 gcgctgatcg agcagctgtt ggaggccatg ctcgccgcgc cgaccgcctc caaccggcag   240
36 gcgtggtcgt tcatggtggt gcgcaggccc gccgcggtcc gccggctgcg cgcgttctcg   300
38 cccggggtgc tgggaacccc cgccttcttc gtcgtggcct gcgtcgaccg cagtctgacc   360
40 gacaacctct ccccgaagct ctcgcagaag atctacgaca ccagcaagct ctgtgtcgcc   420
42 atggcggtgg agaacctgct gctcgcggcg cacgcggccg gcctgggcgg atgcccggtg   480
44 ggcagcttca ggtccgacat cgtcaccagc atgctcggta tcccggaaca catcgagccg   540
46 atgctcgtgg tcccgatcgg ccgtcccgcg acagccctcg tcccctccca gcgccgcgcc   600
48 aagaatgagg tcgtcaacta tgaatcctgg ggaaaccgtg ctgccgcccc aactgcg      657
51 <210> SEQ ID NO: 2
52 <211> LENGTH: 318
53 <212> TYPE: DNA
54 <213> ORGANISM: Streptomyces noursei
56 <400> SEQUENCE: 2
57 atgaatcctg gggaaaccgt gctgccgccc caactgcgtg aggagatcgc gctcctcgcc    60
59 gtctatctgc tcagcagcgg ccgcggactc ctggaggagc cggccgacta cggaatttac   120
61 cgctgtaccg acggggcccg tcgggcgctc caactcctcg acgaacacgg cgggagcacg   180
63 gcacggctga ccgccgtccg cgagcgtctc gacgaggtca tgttcgcgcc gatgggcgag   240
65 gaccgggaca tgggcgcgat tctggacgac ctgtgtcgcc aaatggcaga cgctcttccg   300
67 gaaattgaaa cccccctga                                                318
70 <210> SEQ ID NO: 3
71 <211> LENGTH: 720
72 <212> TYPE: DNA
```

```
73  <213> ORGANISM: Streptomyces noursei
75  <400> SEQUENCE: 3
76  atgcttgcag gcttagttcc cgcgccggac cacggaatgc gggaagaaat acttggcgac   60
78  cgcagccgat tgatccggca acgcggtgag cacgccctca tcggaatcag tgcgggcaac  120
80  agttatttca gccagaagaa caccgtcatg ctgctgcaat gggccgggca gcgtttcgag  180
82  cgcaccgatg tcgtctatgt cgacacccac atcgacgaga tgctgatcgc cgacggccgc  240
84  agcgcgcagg aggccgagcg gtcggtcaaa cgcacgctca aggatctgcg gcgcagactc  300
86  cggcgctcgc tggagagcgt gggcgaccac gccgagcggt tccgtgtccg gtccctgtcc  360
88  gagctccagg agacccctga gtaccgggcc gtacgcgagc gcaccgaccg ggccttcgag  420
90  gaggacgccg aattcgccac cgcctgcgag gacatggtgc gggccgtggt gatgaaccgg  480
92  cccggtgacg gcgtcggcat ctccgcggaa cacctgcggg ccggtctgaa ctacgtgctg  540
94  gccgaggccc cgctcttcgc ggactcgccc ggagtcttct ccgtcccctc ctcggtgctc  600
96  tgctaccaca tcgacacccc gatcacggcg ttcctgtccc ggcgcgagac cggtttccgg  660
98  gcggccgagg acaggcgta cgtcgtcgtc aggccccagg agctggccga cgcggcctag  720
101 <210> SEQ ID NO: 4
102 <211> LENGTH: 834
103 <212> TYPE: DNA
104 <213> ORGANISM: Streptomyces noursei
106 <400> SEQUENCE: 4
107 atgtcatggg gaggacagga cacttgctca tggtgcggaa cggggcccct cggcgaagct   60
109 gaagacgtag gaagacagca cacgtcgcac gccgggggac ccgtcatgac tcaagccgcc  120
111 accgtcaccg ccaccacgag ccagggcagg gcactcctgc ggagcctgac gccgctgttc  180
113 gtggacgccg cgatcccgct cggctcgtac ttcctcctcg ccgagggctt cggcatgagc  240
115 acggtcgccg cgctggcctg gagcagcgtg gtcccggcgc tgcgcacgat ctggggcctg  300
117 gtccgggagc ggacggtcaa cggcctcgcg ctgctgatcc tcgtcgtcaa cgtggtgggg  360
119 ctggcgacga gcaccctgac cggcgatgcc cggctgatga tggccaagga cagcggcgtc  420
121 agcagcgtcg tcgggatcgc gatcctgctc tcggtgcgcg gccggcgccc gctgatgacc  480
123 gccggactcc ggccctgggt gaccaaggga agcccggagg ggaacgccgc atgggaccgg  540
125 ctgtgggcgc gcagcgcgcg gttccggcaa ctggagcggc gattctcgac ggtctggggg  600
127 agcgccctgc tgatcgagtg cgtggtcaag gtcgtcggtg cgtacgtcct gccggtgcac  660
129 accatggtgt ggctgggcac ggtgctgacg gtggtggcga tcctgctggc catggtggtc  720
131 gcgggcggcg gcagcgccga gccgatggag cggatggtca aggccgaggt cggggccgcc  780
133 ggcgaggccg ccacggcggg gaacgccgag ccggcgccgg ccgccgcggc ctga         834
136 <210> SEQ ID NO: 5
137 <211> LENGTH: 3839
138 <212> TYPE: DNA
139 <213> ORGANISM: Streptomyces noursei
141 <400> SEQUENCE: 5
142 ggatccgtcc cgacgggcgg gaaccggtga ggcgccaccc atcgcattcg ccgtaccgcg   60
144 gcgggtgtga ggtgcgccca aaaagaaggg gattgatgtt agctcacagt tcatctgaat  120
146 cgccgccgga atccttgccg gacgcgtgga cggtcctcaa aacccgtacc gccgtccgca  180
148 attacgcgaa agagccggtc gacgacgcgc tgatcgagca gctgttggag gccatgctcg  240
150 ccgcgccgac cgcctccaac cggcaggcgt ggtcgttcat ggtggtgcgc aggcccgccg  300
152 cggtccgccg gctgcgcgcg ttctcgcccg gggtgctggg aaccccccgcc ttcttcgtcg  360
154 tggcctgcgt cgaccgcagt ctgaccgaca acctctcccc gaagctctcg cagaagatct  420
156 acgacaccag caagctctgt gtcgccatgg cggtggagaa cctgctgctc gcggcgcacg  480
158 cggccggcct gggcggatgc ccggtgggca gcttcaggtc cgacatcgtc accagcatgc  540
160 tcggtatccc ggaacacatc gagccgatgc tcgtggtccc gatcggccgt cccgcgacag  600
162 ccctcgtccc ctcccagcgc cgcgccaaga atgaggtcgt caactatgaa tcctggggaa  660
```

```
164 accgtgctgc cgccccaact gcgtgaggag atcgcgctcc tcgccgtcta tctgctcagc   720
166 agcggccgcg gactcctgga ggagccggcc gactacggaa tttaccgctg taccgacggg   780
168 gcccgtcggg cgctccaact cctcgacgaa cacggcggga gcacggcacg gctgaccgcc   840
170 gtccgcgagc gtctcgacga ggtcatgttc gcgccgatgg gcgaggaccg ggacatgggc   900
172 gcgattctgg acgacctgtg tcgccaaatg gcagacgctc ttccggaaat tgaaaccccc   960
174 tgacggctgt ccggggcaac cccaaaagga cttcttagca tgcttgcagg cttagttccc  1020
176 gcgccggacc acggaatgcg ggaagaaata cttggcgacc gcagccgatt gatccggcaa  1080
178 cgcggtgagc acgccctcat cggaatcagt gcgggcaaca gttatttcag ccagaagaac  1140
180 accgtcatgc tgctgcaatg ggccgggcag cgtttcgagc gcaccgatgt cgtctatgtc  1200
182 gacacccaca tcgacgagat gctgatcgcc gacggccgca gcgcgcagga ggccgagcgg  1260
184 tcggtcaaac gcacgctcaa ggatctgcgg cgcagactcc ggcgctcgct ggagagcgtg  1320
186 ggcgaccacg ccgagcggtt ccgtgtccgg tccctgtccg agctccagga gacccctgag  1380
188 taccgggccg tacgcgagcg caccgaccgg gccttcgagg aggacgccga attcgccacc  1440
190 gcctgcgagg acatggtgcg ggccgtggtg atgaaccggc ccggtgacgg cgtcggcatc  1500
192 tccgcggaac acctgcgggc cggtctgaac tacgtgctgg ccgaggcccc gctcttcgcg  1560
194 gactcgcccg gagtcttctc cgtcccctcc tcggtgctct gctaccacat cgacaccccg  1620
196 atcacggcgt tcctgtcccg gcgcgagacc ggtttccggg cggccgaggg acaggcgtac  1680
198 gtcgtcgtca ggccccagga gctggccgac gcggcctagt tggggcgtc cgcgggcgga  1740
200 cctgcctccc cacccgctcc cggtgccggc gccgggcatg acaaatgtca tggggaggac  1800
202 aggacacttg ctcatggtgc ggaacggggc ccctcggcga agctgaagac gtaggaagac  1860
204 agcacgtc gcacgccggg ggacccgtca tgactcaagc cgccaccgtc accgccacca  1920
206 cgagccaggg cagggcactc ctgcggagcc tgacgccgct gttcgtggac gccgcgatcc  1980
208 cgctcggctc gtacttcctc ctcgccgagg gcttcggcat gagcacggtc gccgcgctgg  2040
210 cctggagcag cgtggtcccg gcgctgcgca cgatctgggg cctggtccgg gagcggacgg  2100
212 tcaacggcct cgcgctgctg atcctcgtcg tcaacgtggt ggggctggcg acgagcaccc  2160
214 tgaccggcga tgcccggctg atgatggcca aggacagcgg cgtcagcagc gtcgtcggga  2220
216 tcgcgatcct gctctcggtg cgcggccggc gcccgctgat gaccgccgga ctccggccct  2280
218 gggtgaccaa gggaagcccg gaggggaacg ccgcatggga ccggctgtgg gcgcgcagcg  2340
220 cgcggttccg gcaactggag cggcgattct cgacggtctg ggggagcgcc ctgctgatcg  2400
222 agtgcgtggt caaggtcgtc ggtgcgtacg tcctgccggt gcacaccatg gtgtggctgg  2460
224 gcacggtgct gacggtggtg gcgatcctgc tggccatggt ggtcgcgggc ggcggcagcg  2520
226 ccgagccgat ggagcggatg gtcaaggccg aggtcggggc cgccggcgag gccgccacgg  2580
228 cggggaacgc cgagccggcg ccggccgccg cggcctgaga ccgcgcggcg ggggagttgg  2640
230 ggaaatcgcg tacaggattg cgcgtcgag caccccgcc ctcgataggg cgggcccccg  2700
232 gcgcatatgg tcggcgatgc gacgggacat cggagccccg cgtcgacggt tcaacggcga  2760
234 tccggacggc acgcggcttt cgtcggccac gaagggaacg gaagtcatgt cgactgttca  2820
236 cactggggtc acgcagagcg gtctcaccgc cgagctggcc tccctgcacg ccgagctcgt  2880
238 ccgtcggaat cccggtgaag cggagttcca ccaggcggcc ctggaggtcc tcgaaacgct  2940
240 ggcaccggtg ctcaccgccc ggccggagtt cgccgacgcc aaggtcctgg agcggatcgt  3000
242 cgagccggag cggcagatca tgttccgcgt gccctggcag gacgactccg gcacgatccg  3060
244 ggtcaaccgc ggcttccggg tggagttcaa cagcgcgctc ggcccctaca agggcggcct  3120
246 gcggttccac gcgtccgtca acctcggcat cgtgaagttc ctcggcttcg agcagatctt  3180
248 caagaacgcc ctgaccgggc tgaacatcgg cggcggcaag ggcggcagcg acttcgaccc  3240
250 gcacggcagg tcggacgccg aggtgatgcg cttctgccag tccttcatga ccgagctgca  3300
252 ccgtcacctg ggcgagcaca ccgacgtgcc ggctggcgac atcggcgtcg gcggccggga  3360
254 gatcggctac ctcttcggcc agtaccggcg gatcaccaac cgctgggagg ccggcgtcct  3420
256 gaccggcaag ggcctggcgt ggggcggctc caaggcccgt acggaggcca ccggttacgg  3480
258 caatgtgctg ttcaccgagg agatgctcaa gcagcgcggc gaggagctgg acggccagca  3540
260 ggtggtggtc tccgggtccg gcaacgtcgc catctacacc atcgagaagg cccaggcgct  3600
```
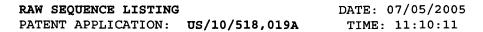
RAW SEQUENCE LISTING   DATE: 07/05/2005
PATENT APPLICATION: US/10/518,019A TIME: 11:10:11

Input Set : A:\PTO.RJ.txt
Output Set: N:\CRF4\07052005\J518019A.raw

```
262 cggcgccaac gtcctgaccg tctcggactc cggcggctac gtcgtcgacg agaagggcat    3660
264 cgacctggcg ctgctcaagc aggtcaagga ggtcgagcgc ggccgggtcg gcgactacgc    3720
266 ccagcggcgc ggcagttcgg cgaagtacgt cgccggcggg agcgtgtggg acgtcgcctg    3780
268 tgacgtggcg ctgccgtcgg ccacccagaa cgagctcgac gcggacgccg cccggatcc     3839
271 <210> SEQ ID NO: 6
272 <211> LENGTH: 219
273 <212> TYPE: PRT
274 <213> ORGANISM: Streptomyces noursei
276 <400> SEQUENCE: 6
278 Met Arg Arg His Pro Ser His Ser Pro Tyr Arg Gly Gly Cys Glu Val
279 1               5                   10                  15
282 Arg Pro Lys Arg Arg Gly Leu Met Leu Ala His Ser Ser Ser Glu Ser
283             20                  25                  30
286 Pro Pro Glu Ser Leu Pro Asp Ala Trp Thr Val Leu Lys Thr Arg Thr
287         35                  40                  45
290 Ala Val Arg Asn Tyr Ala Lys Glu Pro Val Asp Asp Ala Leu Ile Glu
291     50                  55                  60
294 Gln Leu Leu Glu Ala Met Leu Ala Ala Pro Thr Ala Ser Asn Arg Gln
295 65                  70                  75                  80
298 Ala Trp Ser Phe Met Val Val Arg Arg Pro Ala Ala Val Arg Arg Leu
299             85                  90                  95
302 Arg Ala Phe Ser Pro Gly Val Leu Gly Thr Pro Ala Phe Phe Val Val
303             100                 105                 110
306 Ala Cys Val Asp Arg Ser Leu Thr Asp Asn Leu Ser Pro Lys Leu Ser
307         115                 120                 125
310 Gln Lys Ile Tyr Asp Thr Ser Lys Leu Cys Val Ala Met Ala Val Glu
311     130                 135                 140
314 Asn Leu Leu Leu Ala Ala His Ala Ala Gly Leu Gly Gly Cys Pro Val
315 145                 150                 155                 160
318 Gly Ser Phe Arg Ser Asp Ile Val Thr Ser Met Leu Gly Ile Pro Glu
319             165                 170                 175
322 His Ile Glu Pro Met Leu Val Val Pro Ile Gly Arg Pro Ala Thr Ala
323         180                 185                 190
326 Leu Val Pro Ser Gln Arg Arg Ala Lys Asn Glu Val Val Asn Tyr Glu
327         195                 200                 205
330 Ser Trp Gly Asn Arg Ala Ala Ala Pro Thr Ala
331     210                 215
334 <210> SEQ ID NO: 7
335 <211> LENGTH: 104
336 <212> TYPE: PRT
337 <213> ORGANISM: Streptomyces noursei
339 <400> SEQUENCE: 7
341 Asn Pro Gly Glu Thr Val Leu Pro Pro Gln Leu Arg Glu Glu Ile Ala
342 1               5                   10                  15
345 Leu Leu Ala Val Tyr Leu Leu Ser Ser Gly Arg Gly Leu Leu Glu Glu
346             20                  25                  30
349 Pro Ala Asp Tyr Gly Ile Tyr Arg Cys Thr Asp Gly Ala Arg Arg Ala
350         35                  40                  45
353 Leu Gln Leu Leu Asp Glu His Gly Gly Ser Thr Ala Arg Leu Thr Ala
```

RAW SEQUENCE LISTING   DATE: 07/05/2005
PATENT APPLICATION: US/10/518,019A   TIME: 11:10:11

Input Set : A:\PTO.RJ.txt
Output Set: N:\CRF4\07052005\J518019A.raw

```
354      50                 55                 60
357 Val Arg Glu Arg Leu Asp Glu Val Met Phe Ala Pro Met Gly Glu Asp
358 65                 70                 75                 80
361 Arg Asp Met Gly Ala Ile Leu Asp Asp Leu Cys Arg Gln Met Ala Asp
362                 85                 90                 95
365 Ala Leu Pro Glu Ile Glu Thr Pro
366                 100
369 <210> SEQ ID NO: 8
370 <211> LENGTH: 99
371 <212> TYPE: PRT
372 <213> ORGANISM: Streptomyces noursei
374 <400> SEQUENCE: 8
376 Met Leu Pro Pro Gln Leu Arg Glu Glu Ile Ala Leu Leu Ala Val Tyr
377 1                 5                 10                 15
380 Leu Leu Ser Ser Gly Arg Gly Leu Leu Glu Glu Pro Ala Asp Tyr Gly
381                 20                 25                 30
384 Ile Tyr Arg Cys Thr Asp Gly Ala Arg Arg Ala Leu Gln Leu Leu Asp
385         35                 40                 45
388 Glu His Gly Gly Ser Thr Ala Arg Leu Thr Ala Val Arg Glu Arg Leu
389       50                 55                 60
392 Asp Glu Val Met Phe Ala Pro Met Gly Glu Asp Arg Asp Met Gly Ala
393 65                 70                 75                 80
396 Ile Leu Asp Asp Leu Cys Arg Gln Met Ala Asp Ala Leu Pro Glu Ile
397                 85                 90                 95
400 Glu Thr Pro
404 <210> SEQ ID NO: 9
405 <211> LENGTH: 239
406 <212> TYPE: PRT
407 <213> ORGANISM: Streptomyces noursei
409 <400> SEQUENCE: 9
411 Met Leu Ala Gly Leu Val Pro Ala Pro Asp His Gly Met Arg Glu Glu
412 1                 5                 10                 15
415 Ile Leu Gly Asp Arg Ser Arg Leu Ile Arg Gln Arg Gly Glu His Ala
416                 20                 25                 30
419 Leu Ile Gly Ile Ser Ala Gly Asn Ser Tyr Phe Ser Gln Lys Asn Thr
420         35                 40                 45
423 Val Met Leu Leu Gln Trp Ala Gly Gln Arg Phe Glu Arg Thr Asp Val
424       50                 55                 60
427 Val Tyr Val Asp Thr His Ile Asp Glu Met Leu Ile Ala Asp Gly Arg
428 65                 70                 75                 80
431 Ser Ala Gln Glu Ala Glu Arg Ser Val Lys Arg Thr Leu Lys Asp Leu
432                 85                 90                 95
435 Arg Arg Arg Leu Arg Arg Ser Leu Glu Ser Val Gly Asp His Ala Glu
436                 100                 105                 110
439 Arg Phe Arg Val Arg Ser Leu Ser Glu Leu Gln Glu Thr Pro Glu Tyr
440         115                 120                 125
443 Arg Ala Val Arg Glu Arg Thr Asp Arg Ala Phe Glu Glu Asp Ala Glu
444       130                 135                 140
447 Phe Ala Thr Ala Cys Glu Asp Met Val Arg Ala Val Val Met Asn Arg
```

RAW SEQUENCE LISTING ERROR SUMMARY   DATE: 07/05/2005
PATENT APPLICATION: US/10/518,019A   TIME: 11:10:12

Input Set : **A:\PTO.RJ.txt**
Output Set: **N:\CRF4\07052005\J518019A.raw**

<u>Please Note:</u>
Use of n and/or Xaa have been detected in the Sequence Listing. Please review the
Sequence Listing to ensure that a corresponding explanation is presented in the <220>
to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:12; Xaa Pos. 8
Seq#:13; Xaa Pos. 3

<u>Invalid <213> Response:</u>
Use of "Artificial" only as "<213> Organism" response is incomplete,
per 1.823(b) of New Sequence Rules. Valid response is Artificial Sequence.

Seq#:14,15,16,17,18,19,20,21,22,23

VERIFICATION SUMMARY
PATENT APPLICATION: **US/10/518,019A**

DATE: 07/05/2005
TIME: 11:10:12

· Input Set : **A:\PTO.RJ.txt**
 Output Set: **N:\CRF4\07052005\J518019A.raw**

L:578 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:12 after pos.:0
L:595 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:13 after pos.:0